Differential Cryptanalysis

Vincent Rijmen

Thanks to Stefan Mangard, Joan Daemen

Outline

- Basic
  - Differential Cryptanalysis of 3 rounds of DES
  - Differential Cryptanalysis of 6 rounds of DES
- Advanced
  - Probability
  - Differentials and characteristics
  - Markov ciphers
  - Decorrelation theory
  - Key-alternating ciphers
- AES (Wide trail design strategy)

Differential cryptanalysis

- Biham and Shamir, 1991
- Chosen plaintext attack
- Goal: Determine the secret key
- Exploits mapping properties of differences within DES

Difference propagation

- Notation:
  - Pair of values X, X^*
  - Difference $X' = X^* - X = X^* + X$
- Linear map L, by definition:

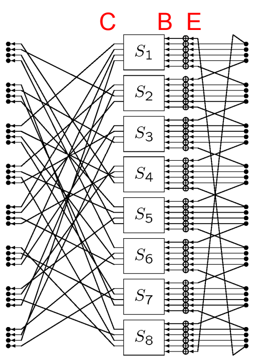
$$L(X^*) + L(X) = L(X^* + X) = L(X')$$

- Addition with constant (key)

$$(X^* + K) + (X + K) = X^* + X = X'$$

- Nonlinear map S
  - If $X^* == X$ then $S(X^*) - S(X) = 0$
  - Else $S(X^*) - S(X) = ?$

The DES “F” function: notation



The set of possible inputs to an Sbox

- The set of possible inputs for given input- and output differences for Sbox j:

$$IN(B_j', C_j') = \{ B_j : S_j(B_j) + S_j(B_j + B_j') = C_j' \}$$

- Important observation: Not every input difference can produce every output difference

The set of possible inputs to an Sbox

S_1

14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

Example:
IN(110100,0100) = {010011,100111}

$IN(B_j', C_j') = \{ B_j : S_j(B_j) + S_j(B_j + B_j') = C_j' \}$

The key XOR

$B' = B + B^* = (E + K) + (E^* + K) = E + E^* = E'$

- The input difference of the Sboxes of a round does not depend on the round key
- Important observation: $E_j + K_j \in IN(E_j', C_j')$

The set of all keys that are possible

- The set of possible input values

$IN(B_j', C_j') = \{ B_j : S_j(B_j) + S_j(B_j + B_j') = C_j' \}$

- The set of possible keys:

$Test_j(E_j, E_j^*, C_j') = \{ E_j + B_j : B_j \in IN(E_j', C_j') \}$

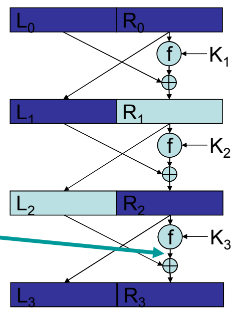
$K_j = B_j + E_j$

- Given E, E^* and C', we can narrow down the key space

Attack on 3 rounds of DES

E and E^* of the last round are known!

The important question is: What is C´?



Attack on 3 rounds of DES

$R_3 = L_0 + f(R_0,K_1) + f(R_2,K_3)$

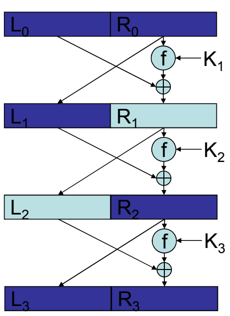
The differences can be expressed as follows, if $R_0' = 0$:

$R_3' = L_0' + f(R_2,K_3) + f(R_2^*,K_3)$

C'

The set of possible values for K_3 can be determined:

$test_j(E_j, E_j^*, C_j')$



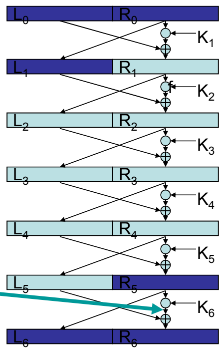
Attack on 3 rounds of DES

- Last round key can be determined by using several pairs of plaintexts with $R_0' = 0$:
  - Key is in intersection of the sets of possible values
- Then we know 48 bits of the 56-bit key
- Remaining 8 key bits: try out all possibilities

Attack on 6 rounds of DES

- C' cannot be calculated as easily as before
- Hence, a probabilistic approach is pursued

What is C´?

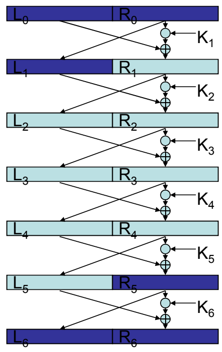


Attack on 6 rounds of DES

- Definition of a characteristic:

  L'_0, R'_0
  L'_1, R'_1, p_1
  L'_2, R'_2, p_2
  …
  L'_n, R'_n, p_n

- p_i is the *probability* that L'_{i-1}, R'_{i-1} is mapped to L'_i, R'_i



Probability in differential cryptanalysis

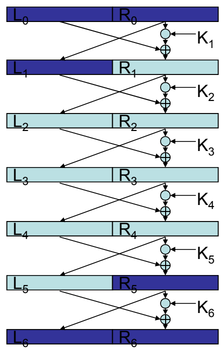
- Frequentist definition: probability denotes the *relative frequency of occurrence* of a certain outcome of an *experiment*, when *repeating* the experiment.
- Experiment: encrypt 1 pair of plaintexts under 1 key
- Repeating: different plaintexts and/or different key
- Standard description of the differential attack assumes: different plaintexts, same key
- Most theory assumes: different key
- Implicit ergodicity assumption

Attack on 6 rounds of DES

- 1-round characteristic:

L'_0 = anything	R'_0 = 00000000
L'_1 = 00000000	$R'_1 = L'_0$

$p_1 = 1$

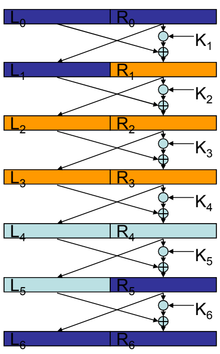


Attack on 6 rounds of DES

- 3-round characteristic:

L'_0 = 40080000	R'_0 = 04000000
L'_1 = 04000000	R'_1 = 00000000
L'_2 = 00000000	R'_2 = 04000000
L'_3 = 04000000	R'_3 = 40080000

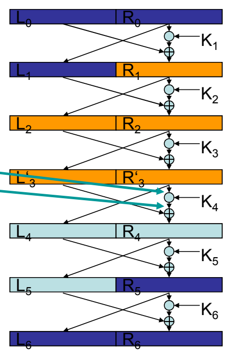
$p_1 = 0.25$
$p_2 = 1$
$p_3 = 0.25$



Attack on 6 rounds of DES

With probability 1/16:
L'_3 = 04000000 R'_3 = 40080000

In that case:
Input and Output difference for S_3,S_5,S_6,S_7 and S_8 = 0



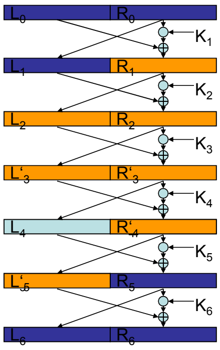
Attack on 6 rounds of DES

$R_6' = C' + L_5'$

$R_6' = C' + L_3' + f(K_4, R_3) + F(K_4, R_3^*)$

We can compute C' for the 5 S-boxes where $R_3' = 0$

The keys J_3, J_5, J_6, J_7 and J_8 can be determined!



Wrong pairs

- 15 out of 16 times, the pair doesn't follow the characteristic
- 10 out of these 15 times we get at least one empty $test_i$
- We can *filter* this pair
- 5/15 of the wrong pairs can't be filtered ⇒ random key suggestions = *noise*

- Keys in test set are *suggested* keys
- After some time the right key should be among the most suggested values

Signal-to-noise ratio

- Let α = average number of keys in test set
- β = fraction of unfiltered wrong pairs
- 2^k = number of keys

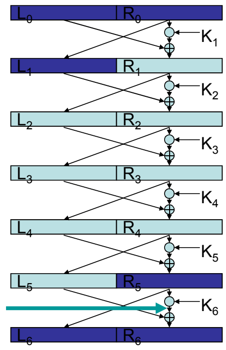
$$S/N = p/(\alpha\beta / 2^k) = 2^k\, p/(\alpha\beta)$$

- We need at least 2/p pairs to discover the right key
- Make k as large as possible (memory constraints)

Summary of the attack

- It is necessary to determine the output differences of the Sboxes in the last round

- A "good" characteristic needs to be found in order to get there



Security against differential attacks

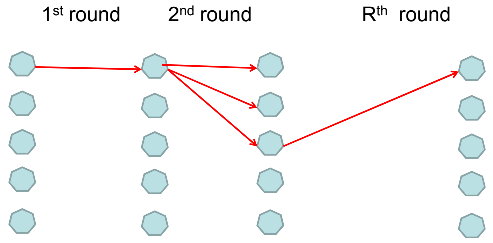
- Make prediction of differences difficult
- Ensure that there are no high-probability characteristics
  - Compute bounds for existing ciphers
  - Design ciphers with low bounds on the probability
  - Design ciphers with easily computable bounds

Computing bounds for DES

- Done by determining the best characteristics
- A* algorithm: branch and prune, depth-first
- Determine iteratively the best characteristic over 1, 2, 3, … rounds

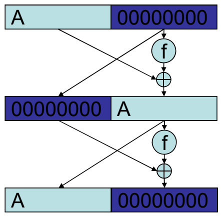
- Prune: if cost of current path over t rounds + cost of best path over (R-t)-rounds ≥ cost of currently best path over R rounds, then abandon the current path

A* algorithm



Results for DES

- The best characteristics over 8 rounds or more, are iterative characteristics
- Two values for A possible
- With 3 *active* S-boxes
- Probability = 1/234 for every two rounds



Differential strengthening of DES

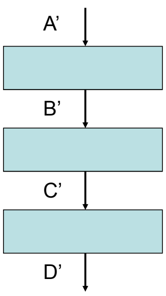
- The S-box design criteria (+ expansion) ensure that iterative characteristics have at least 3 active S-boxes
- Any re-ordering of the S-boxes would increase the probability of the best characteristic
- DES designers knew about differential cryptanalysis
- On the other hand, it is possible to find S-boxes that behave better in this respect

Technical problems

Computing the probability

1. Characteristics and differentials
2. Independence of rounds

Predicting a difference



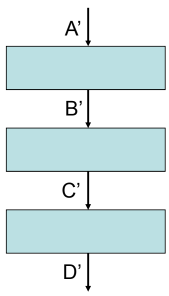
$P_1 = \Pr(A' \to B')$

$P_2 = \Pr(B' \to C')$

$P_3 = \Pr(C' \to D')$

$\Pr(A' \to D') = p_1 p_2 p_3$???

Characteristics and differentials



$\Pr(A' \to D') =$

$\Pr(A' \to B' \to C' \to D')$

$+ \Pr(A' \to B_1' \to C_1' \to D')$

$+ \ldots$

$= \Sigma_{B'} \Sigma_{C'} \Pr(A' \to B' \to C' \to D')$

(A',D'): differential

(A',B',C',D'): characteristic (trail, path)

Characteristic and differential probabilities

- Pr(A',D') ≥ Pr(A',B',C',D')
- Computing Pr(A',D') is more difficult than computing Pr(A',B',C',D')
- In a 'weak' cipher, usually one characteristic dominates the probability: Pr(A',D') ≈ Pr(A',B',C',D')
  - In many 'strong' ciphers: open problem

Computing Pr(A' → B' → C' → D')

- Pr(A' → B') × Pr(B' → C') × Pr(C' → D') ??
- Actually:
  Pr(A' → B') × Pr(B' → C' | A') × Pr(C' → D' | A', B')
- Theory of Markov ciphers [Lai,Massey,Murphy]

Markov cipher

- Definition: cipher such that over one round:
  Pr(A' → B') = Pr(A' → B' | X)
- With X: input value
  - Obviously, Pr here is computed over different keys
- Definition of EDP:
  EDP(A'→B'→C'→D') = Pr(A'→B') × Pr(B'→C') × Pr(C'→D')
- Fundamental Theorem: EDP(A'→B'→C'→D') equals 'probability' if all rounds use independent keys.

Hypothesis of stochastic equivalence

- EDP ≈ E[Pr(A'→B'→C'→D')]
  - Given 1 pair with input difference A', the probability that it has differences B', C', and D'
- Related quantity: DP[k]
  - Given q pairs with input difference A', the fraction that will have differences, B', C', D'
  - *Probability* computed with fixed key
- Hypothesis [Lai,Massey,Murphy]:
  For almost all keys k:
  DP[k](A'→B'→C'→D') ≈ EDP(A'→B'→C'→D')

Problems with the hypothesis of S.E.

1. Computing EDP of a differential remains a problem
2. The hypothesis doesn't hold
   - Example: DES:
   - Probability of the best characteristic: 2 rounds EDP = 1/234
   - For 13 rounds (used in attack), EDP = 2^{-47}
   - 2 rounds DP[k] = 1/146 or 1/585
   - For 13 rounds, this gives $2^{-43} \leq$ DP[k] $\leq 2^{-55}$

Hypothesis of S.E. can't hold

- DP[k](A'→B'→C'→D') is always a multiple of (No. of pairs)$^{-1}$
- EDP can become much smaller:
  (No. of pairs)$^{-1}$ × (No. of keys)$^{-1}$
- For modern ciphers, EDP < (No. of pairs)$^{-1}$
  - Impact on DP[k] ???
- Nevertheless, we continue with EDP

Provable security (Knudsen/Nyberg)

- Developed for Feistel ciphers
- Prove upper bounds on the EDP of a differential through the cipher

Theorem:

If for 2 rounds EDP(A',D') $\leq$ p

Then for 4 or more rounds EDP(A',D') $\leq 2p^2$

- Extension: $\leq p^2$ if f-function is bijective
- Examples: Misty, KASUMI

- Problem: doesn't improve after 4 rounds

Decorrelation theory (Vaudenay)

- Borrows techniques from universal hash function design

- Example: $F(X,K) = K_1 \times X + K_2$
  - $F(X, K) + F(X+A',K) = (K_1 \times X + K_2) + (K_1 \times (X+A') + K_2)$
    $= A' \times K_1$
  - $DP[k](A' \rightarrow B') = 1$ if $B' = A' \times K_1$
    $= 0$ otherwise
  - $EDP(A' \rightarrow B') = (\text{No. of keys})^{-1}$

- Very good bound on EDP

Attack

- Example: $F(X,K) = K_1 \times X + K_2$
- Consider X, X+A', X+B', X+A'+B'

F(X, K) + F(X+A', K) + F(X+B', K) + F(X+A'+B',K) =

$A' \times K_1 + A' \times K_1 = 0$

- Characteristic with EDP 1!

- Demonstrates problem of this notion of provable security

Wide trail design strategy

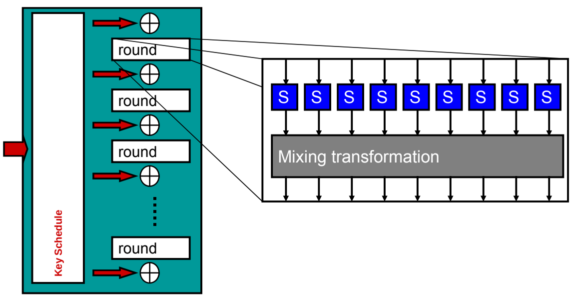
- Compute bounds for 1 S-box:
  $d = \max_{A' \neq 0, B'} \Pr(A' \rightarrow B')$

- Compute bound on number of *active* S-boxes
  z = minimum number of active S-boxes

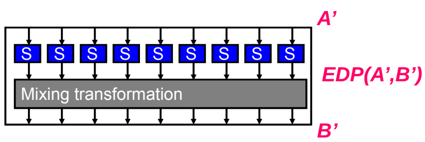
- Together: EDP $\leq d^z$

- Bound valid for characteristics, not differentials

Iterative cipher



Single-Round Optimization



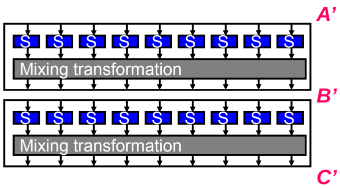
Relevant:

- Number of active components in *A*
- Worst-case difference propagation probability in S-box

Provides a bound of 1 active S-box per round

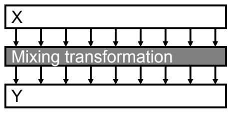
➢ Small d $\Rightarrow$ Low bound requires large S-boxes

Two-Round Optimization

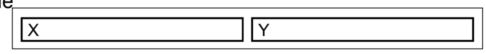


- Relevant: number of active components in (*A'*,*B'*)
- Diffusion criterion for mixing transformation *y* = m(*x*)
  - Branch number $\mathcal{B}$: minimum number of active comp. in (*A'*,*B'*)
- $\mathcal{B}$ depends only on the mixing transformation

Designing the Mixing Transformation



- $\mathcal{B} \leq$ number of components of X plus 1
- (x,y) with y = m(x) can be seen as an error-correcting code
- $\mathcal{B}$ corresponds with the minimum distance of this code
- Maximum $\mathcal{B}$: take a Maximum Distance Separable (MDS) code

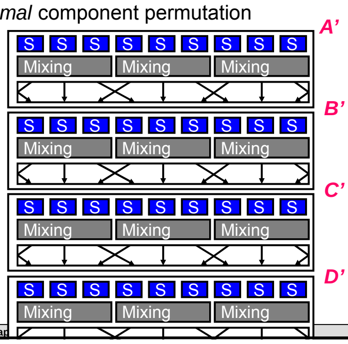


Shark

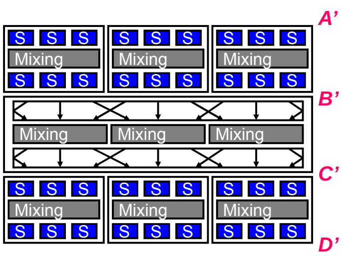
- Block length of 64 bits = 8 bytes
- 8-bit S-box
- MDS code over GF(256), length 16, dimension 8
- Optimal 2-round mixing
- Sub-optimal performance

Four-Round Optimization (1)

- Compose linear part of local mixing transformation and *diffusion optimal* component permutation



Four-Round Optimization (2)



- Reorder transformations ⇒ *Super-boxes*
- Apply two-round theorem recursively: $\mathcal{B}^2$ active S-boxes

Square

- Block length of 128 bits = 16 bytes = 4 × 4
- 8-bit S-box
- MDS code over GF(256), length 8, dimension 4
- Diffusion optimal permutation: transpose
- 4-round mixing: 25 active S-boxes per 4 rounds
- S-box: EDP $\leq 2^{-6}$
- EDP of 4-round characteristic $\leq 2^{-150}$

Rijndael

- Preliminary AES call asked for variable block length
  - Needed rectangular input arrays
  - Replace transpose by row shift
- Increase number of rounds (improved cryptanalysis)

- PR
  - More complicated key schedule
  - Use ObjectOriented names for different components

Remark

- MDS codes require byte-level approach

- Similar approach, but on bit level, by Tavares et al. [1998]
- Diffusion on bit level
  - Also within the S-boxes (Avalanche criteria)

Conclusions

- Differential cryptanalysis
  - Basic method
  - Several theories to secure designs
  - Simple AES structure allows for easier computation of bounds